

## 平成25年度科学研究費助成事業（特別推進研究）自己評価書 〔追跡評価用〕

◆記入に当たっては、「平成25年度科学研究費助成事業（特別推進研究）自己評価書等記入要領」を参照してください。

平成25年 4月 25日現在

<b>研究代表者 氏名</b>	有村 博紀	<b>所属研究機関・ 部局・職</b>	北海道大学・大学院情報科学研究科・教授
<b>研究課題名</b>	知識基盤形成のための大規模半構造データからの超高速パターン発見		
<b>課題番号</b>	17002008		
<b>研究組織 (研究期間終了時)</b>	研究代表者 有村 博紀（北海道大学・大学院情報科学研究科・教授） 研究分担者 喜田 拓也（北海道大学・大学院情報科学研究科・准教授） 湊 真一（北海道大学・大学院情報科学研究科・准教授） 伊藤 公人（北海道大学・人獣共通感染症リサーチセンター・准教授）		

### 【補助金交付額】

年度	直接経費
平成17年度	43,800 千円
平成18年度	64,178 千円
平成19年度	7,300 千円
総計	115,278 千円

## 1. 特別推進研究の研究期間終了後、研究代表者自身の研究がどのように発展したか

特別推進研究によってなされた研究が、どのように発展しているか、次の(1)～(4)の項目ごとに具体的かつ明確に記述してください。

### (1) 研究の概要

(研究期間終了後における研究の実施状況及び研究の発展過程がわかるような具体的内容を記述してください。)

本研究期間中は、ネットワーク上に蓄積された大規模半構造データからの知識基盤形成のための基礎技術である大規模半構造データマイニング技術の研究を行った。本研究期間中は、次の項目の研究を実施し、いずれも確固たる成果を獲得した。

#### (1) 最適パターン発見に基づく超高速半構造マイニング手法の研究開発。

申請者らが開発した半構造マイニング技術に基づき、木構造や、系列、時空間データ等の大規模半構造データに対して、世界で初めての性能保証をもつ超高速かつ低メモリのパターンマイニング手法を確立した。

#### (2) 極大パターン発見に基づく超高速半構造マイニング手法の研究開発。

解の総数を著しく減少させながら、情報損失のないパターン発見を行う極大パターン発見の理論を解明し、これに基づいて、さまざまな大規模半構造データに対する高速な極大パターン発見手法を開発した。

#### (3) 大規模知識基盤形成システムのための半自動的知的関係技術の研究開発。

周辺技術として、大規模非構造ストリームデータに対する高速な照合と圧縮の技術を開発した。

#### (4) 大規模知識基盤形成システムのための高速な知識索引技術の研究開発。

分担者の一人(湊)が開発したゼロサプレス二分決定グラフ技術に基づき、大規模半構造データを圧縮すると同時に、自由に操作するための知識索引技術を開発した。

期間終了後は、本研究の成果を基盤として、下記の課題の研究開発を実施した。

#### ● (1) 最適パターン発見と極大パターン発見に基づく超高速半構造マイニング手法の研究開発。

- ・高速イベントストリームデータからの、エピソード発見と呼ばれる新種の時系列パターン発見問題に取り組み、性能保証をもつ高速なアルゴリズムを開発した。

- ・誤差やノイズを許したあいまい頻出集合発見問題に対する性能保証をもつ高速なアルゴリズムを開発した。

- ・実世界の連続的時空間を扱う大規模時空間データからの理論的性能保証をもつ極大パターン発見アルゴリズムの研究開発に成功した。

- ・種々の離散構造に対する極大パターン発見に基づく超高速半構造マイニングの機構を理論的に解明し、高速アルゴリズムの一般的な設計理論を与えた。

#### ● (2) 大規模知識基盤形成システムのための半自動的知的関係技術の研究開発。

- ・ハードウェア指向の半自動的知的関係技術の新展開。ワード内 SIMD および簡潔データ構造と呼ばれる最先端のアルゴリズム技術に基づき、超高速半構造ストリームに対して、パターン照合アルゴリズムをハードウェア上で直接実行可能な超高速アルゴリズム設計法を開発した。

#### ● (3) 高速な知識索引技術の研究開発の新展開。

- ・ゼロサプレス二分決定グラフ技術を、系列ストリームデータや巡回データ等の半構造データへ拡張し、これらの半構造データに対する知識索引の実現方法を明らかにした。

- ・ゼロサプレス二分決定グラフの演算計算量に関して、1980年代の同技術の提案時から広く信じられてきた予想の反例を与え、予想を否定的に解決して決着をつけた。一方で、系列データに対して、開発した知識索引が接尾辞索引構築などのいくつかの問題を効率よく解くことができることを示し、半構造データに対する知識索引の効率と限界を明らかにした。

## 1. 特別推進研究の研究期間終了後、研究代表者自身の研究がどのように発展したか（続き）

(2) 論文発表、国際会議等への招待講演における発表など（研究の発展過程でなされた研究成果の発表状況を記述してください。）

### ● 招待講演

- [1] Hiroki Arimura: Efficient Algorithms for Mining Frequent and Closed Patterns from Semi-structured Data, Proc. 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2008), Osaka, 2008. (基調講演)
- [2] Hiroki Arimura: 25th International Symposium on Computer and Information Sciences (ISCIS'10), London, 2010.(招待講演)
- [3] Hiroki Arimura: International Workshop on Statistical-Mechanical Informatics 2010 (IW-SMI 2010), DEX-SMI, The Physical Society of Japan, IEICE Engineering Sciences Society, Shiran-Kaikan, Kyoto, Japan, March 7-10, 2010 .(招待講演)

### ● (1) 最適パターン発見発見に基づく超高速半構造マイニング手法の研究開発.

#### ◎エピソードマイニング

- [4] Takashi Katoh, Hiroki Arimura and Kouichi Hirata: A Polynomial-delay Polynomial-space Algorithm for Extracting Frequent Diamond Episodes from Event Sequences, Lecture Notes in Computer Science 5476, Proc. 13rd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2009), Springer, 172-183, 2009.
- [5] Takashi Katoh, Hiroki Arimura and Kouichi Hirata: Mining Frequent Bipartite Episodes from Event Sequences, Lecture Notes in Computer Science 4265, Proc. 12th International Conference on Discovery Science (DS2009), Springer, 136-151, 2009.
- [6] Takashi Katoh, Hiroki Arimura and Kouichi Hirata: Mining Frequent k-Partite Episodes from Event Sequences, Lecture Notes in Computer Science 6284, New Frontiers in Artificial Intelligence, Springer, 331-344, 2009.

#### ◎超高速木構造マイニング

- [7] Kunihiro Wasa, Takeaki Uno, and Hiroki Arimura: Constant Time Enumeration of Bounded-Size Subtrees in Trees and Its Application, Lecture Notes in Computer Science 7434, Proc. 18th Annual International Computing and Combinatorics Conference (COCOON'12), Springer, 2012 (to appear).

### ● (2) 極大パターン発見に基づく超高速半構造マイニング手法の研究開発.

#### ◎極大パターン発見

- [8] Hiroki Arimura and Takeaki Uno: Mining Maximal Flexible Patterns in a Sequence, New Frontiers in Artificial Intelligence, *Lecture Notes in Computer Science* 4914, Springer, 307-317, 2008.
- [9] Takeaki Uno and Hiroki Arimura: Ambiguous Frequent Itemset Mining and Polynomial Delay Enumeration, *Lecture Notes in Computer Science* 5012, Proc. 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2008), Springer, 2008.

#### [10] ◎時空間パターンマイニング)

- [11] H. Arimura, T. Uno and S. Shimosono: Time and Space Efficient Discovery of Maximal Geometric Graphs, *Lecture Notes in Computer Science* 4755, Proc. the 10th International Conference on Discovery Science (DS-2007), Springer, 42-55, 2007.
- [12] X. Geng, H. Arimura, T. Uno: Pattern Mining from Trajectory GPS Data, Proc. the 3rd IIAI International Conference on e-Services and Knowledge Management (IIAI ESKM'12), 2012.

#### ◎半構造データマイニングの一般理論の構築

- [13] H. Arimura and T. Uno: Polynomial-Delay and Polynomial-Space Algorithms for Mining Closed Sequences, Graphs, and Pictures in Accessible Set Systems, Proc. 2009 SIAM International Conference on Data Mining 2002 (SDM'09), SIAM, 1087-1098, 2009.

### ● (3) ハードウェア指向の半自動的知的連係技術の新展開.

### ◎イベントストリーム処理

- [14] Tomoya Saito, Takuya Kida, Hiroki Arimura: An Efficient Algorithm for Complex Pattern Matching over Continuous Data Streams Based on Bit-Parallel Method, The Third IEEE International Workshop on Databases for Next-Generation Researchers (SWOD'07), IEEE, 13-18, 2007.
- [15] Yusaku Kaneta and Hiroki Arimura: Faster Bit-Parallel Algorithms for Unordered Pseudo-tree Matching and Tree Homeomorphism, Lecture Notes in Computer Science 6460, Proc. 21st International Workshop on Combinatorial Algorithms (IWOC2010), 68-81, March 2010.
- [16] Yusaku Kaneta, Shin-ichi Minato, and Hiroki Arimura: Fast Bit-Parallel Matching for Network and Regular Expressions, Lecture Notes in Computer Science 6393, Proc. the 17th Symposium on String Processing and Information Retrieval (SPIRE2010), 372-384, 2010.
- [17] Hirohito Sasakawa, Hiroki Arimura: Trajectory Pattern Matching Based on Bit-Parallelism for Large GPS Data, Proc. the 3rd IIAI International Conference on e-Services and Knowledge Management (IIAI ESKM'12), IIAI, 2012 .
- [18] Yusaku Kaneta, Shingo Yoshizawa, Shin-ichi Minato, Hiroki Arimura, and Yoshikazu Miyanaga: A Dynamically Reconfigurable FPGA-based Pattern Matching Hardware for Subclasses of Regular Expressions, IEICE Trans. Inf. and Syst., Vol.E95-D, No.7, 1847-1857, July 2012.
- [19] Yusaku Kaneta, Hiroki Arimura, and Rajeev Raman: Faster Bit-Parallel Algorithms for Unordered Pseudo-tree Matching and Tree Homeomorphism, Journal of Discrete Algorithms, Vol.14, 119-135, July 2012.

### ◎半構造データ索引

- [20] T. Uemura and H. Arimura: A Linear-Time Off-line Construction of Property Suffix Trees, IEICE Trans. Inf. and Syst., J91-D(3), 595-607, 2008 (in Japanese).
- [21] Takashi Uemura, Daisuke Ikeda, Takuya Kida and Hiroki Arimura: Unsupervised Spam Detection by Document Complexity Estimation with Maximal Overlap Method, Transactions of the Japanese Society for Artificial Intelligence, Vol.26, No.1, 297-306, 2011.
- [22] Takashi Uemura and Hiroki Arimura: Sparse and Truncated Suffix Trees on Variable-Length Codes, Lecture Notes in Computer Science 6661, Proc. the 22nd Annual Symposium on Combinatorial Pattern Matching (CPM2011), Palermo, Springer, 246-260, June 2011.
- [23] Takuya Takagi, Takashi Uemura, and Hiroki Arimura: Speeding Up Compact Trie Structures on Word RAM and Its Applications, Proc. Third Workshop on Algorithms for Large-Scale Information Processing (ALSIP 2012), Technical Note, Dec. 2012.

### ● (4) 高速な知識索引技術の研究開発の新展開.

### ◎知識索引 ZDD と超高速極大集合発見手法 LCM の連携

- [24] Shin-ichi Minato, Hiroki Arimura: Frequent Closed Item Set Mining Based on Zero-suppressed BDDs, Proc. of International Workshop on Data Mining and Statistical Science (DMSS-2006), 77-84, 2006.
- [25] T. Uno and H. Arimura: An Efficient Polynomial Delay Algorithm for Pseudo Frequent Itemset Mining, Lecture Notes in Computer Science 4755, Proc. the 10th International Conference on Discovery Science (DS-2007), Springer, 219-230, 2007.
- [26] S. Minato, T. Uno, and H. Arimura: LCM over ZBDDs: Fast Generation of Very Large-Scale Frequent Itemsets Using a Compact Graph-Based Representation, Lecture Notes in Computer Science 5012, Proc. 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2008), Springer, 2008.
- [27] Shin-ichi Minato and Hiroki Arimura: Frequent Closed Item Set Mining Based on Zero-suppressed BDDs, Transactions of the Japanese Society for Artificial Intelligence, Vol.22, No.2, 165-172, 2007.
- [28] Ryo Yoshinaka, Jun Kawahara, Shuhei Denzumi, Hiroki Arimura, and Shin-ichi Minato: Counterexamples to the long-standing conjecture on the complexity of BDD binary operations, Information Processing Letters, Vol.112, Issue 16, pp.636-640, Aug. 2012.

### ◎知識索引の半構造データへの拡張

- [29] Shuhei Denzumi, Ryo Yoshinaka, Hiroki Arimura and Shin-Ichi Minato: Notes on Sequence Binary Decision Diagrams: Relationship to Acyclic Automata and Complexities of Binary Set Operations, Proc. Prague Stringology Conference 2011 (PSC'11), Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague 2009.

## 1. 特別推進研究の研究期間終了後、研究代表者自身の研究がどのように発展したか（続き）

### (3) 研究費の取得状況（研究代表者として取得したもののみ）

◇科研費補助金 基盤研究(A)「大規模知識基盤形成のための次世代半構造マイニング技術の研究」  
(2008-2011 年度, 50,440 千円, 代表 有村):

本特別推進研究の研究成果を基に、超高速半構造マイニング技術と知識索引および知識連携技術を融合して、ウェブ上のネットワークデータからの知識獲得のためのアルゴリズム開発とプロトタイプ構築を行った。とくに、本特別推進研究の成果を、不完全データや、時系列データからの半構造マイニングへ拡張し、高速なアルゴリズムを開発し、一方でその一般理論の構築に成功するなどの成果を上げている。

◇科学研究費補助金 基盤研究(A)「大規模知識基盤形成のための次世代半構造マイニング技術の展開」  
(2012-2015 年度, 40,370 千円, 代表 有村):

上記の基盤研究(A)を発展させ、実世界の大規模非構造情報からの知識基盤形成を支える知識獲得のためのアルゴリズム開発とプロトタイプ構築を行った。特に、実世界のビッグデータ（大規模非構造データ）に対する時間パターン照合やストリーム処理技術、知識索引技術に焦点をあてて研究を推進している。

◇文科省グローバル GCOE プログラム「知の創出を支える次世代 IT 基盤拠点」  
(2007-2011 年度, 北大情報科学研究科, 拠点リーダー 有村):

拠点リーダーとして、本特別推進研究の知識基盤形成のための知識発見技術の着想に基づいて、実世界と情報世界を融合する知の創出のための次世代情報技術推進を構想し、同拠点において異分野共同研究プロジェクト制等の教育研究拠点形成を行った。本プログラムは採択時に重点配分拠点に選定され、また、事後評価において A 評価（十分な成果をあげた）を受けるなど高く評価された。

### (4) 特別推進研究の研究成果を背景に生み出された新たな発見・知見

#### ◎大規模半構造データに対する超高速かつ低メモリのパターンマイニング手法の確立:

最適パターン発見と極大パターン発見において、申請者らが開発した超高速極大集合発見技術 LCM (2004 年) を、木や、系列、グラフ等の種々の半構造データに拡張することに成功し、多項式遅延計算の理論的な性能保証をもつアルゴリズムの統一的な設計原理を明らかにした。このための高速化技術である「接頭辞保存拡張」(PPC-extension) 手法は、その後の種々の半構造マイニングアルゴリズムの基礎となっている。

#### ◎半構造データマイニングにおける知識索引技術の有効性:

・1990 年代に大規模論理回路設計 (VLSI CAD) 分野で広く使われてきた知識索引 (ZDD 索引, ゼロサプレス二分決定グラフ, 分担者 湊) の大量データからの知識発見における有効性を示した。これを実現するために、ZDD ベクターやパターン閉包演算等の知識発見技術が新たに研究開発された。本研究成果は、2006 年に学術論文 (参考論文 [1]) として出版され、その後の技術動向を予見した論文として、2010 年電子情報通信学会 情報・システムソサイエティ 論文賞 (先見論文) を 2010 年 6 月に受賞した。

・知識索引技術の有効性として、ZDD によるパターン閉包演算 (論文 1(2) [27]: 湊・有村) は、情報科学のバイブルと呼ばれる Donald Knuth 博士の名著「The Art of Computer Programming」(Vol.1, Fascicle.1) でも、新しい ZDD 演算として紹介された。

#### ◎知識索引技術と超高速パターン発見技術の融合:

別個に研究されてきたパターン発見 (LCM アルゴリズム) と大規模知識索引 (ZDD 索引) を、有機的に連携させて、大量データから発見した膨大なパターンを陽に表現することなく、直接、圧縮した形で知識索引に格納し、圧縮したままでさまざまな知識発見演算を実行可能であることを示した (LCMoverZDD 技術)。

受賞: 2010 年電子情報通信学会 情報・システムソサイエティ 論文賞 (先見論文), 湊真一, 有村博紀, 平成 21 年 6 月 1 日授賞。(ゼロサプレス型二分決定グラフを用いたトランザクションデータ ベースの効率的解析手法, 湊真一, 有村博紀, 電子情報通信学会論文誌 D, Vol. J89-D, No. 2, 172-182, 2005.)

## 2. 特別推進研究の研究成果が他の研究者により活用された状況

特別推進研究の研究成果が他の研究者に活用された状況について、次の(1)、(2)の項目ごとに具体的かつ明確に記述してください。

### (1) 学界への貢献の状況（学術研究へのインパクト及び関連領域のその後の動向、関連領域への関わり等）

#### ◎半構造マイニングの発展と、その種々の基盤アルゴリズムの確立：

- ・本特別推進研究の中心課題の一つである大規模で非定型なデータからの知識発見（すなわち、「半構造データマイニング」）は、2010年代に入り、いわゆる「ビッグデータからの知識発見」として、学術及び産業界から次世代情報処理技術の中心課題として、一躍注目されるようになった。本特別推進研究は2005年のスタート時から一貫して、これらのビッグデータ（大規模非定型データ）に対する高速処理アルゴリズムの研究開発に取り組み、ビッグデータ研究に関する世界的にも先駆的なプロジェクトである。
- ・これらのビッグデータは、主要な対象として非構造な離散構造データや超高速ストリームデータを想定しており、それらからの高速マイニングがその挑戦的な技術課題の一つであることが、現在、世界中で広く認識されている。本特別推進研究は、これらのネットワーク上の離散構造とストリームデータを対象とした超高速データマイニング技術を中心に2005年から2007年にかけて集中的な研究開発を実施し、2000年代初頭から勃興してきた半構造マイニング技術分野において、申請者と分担者らの提案・開発による「最右拡張技術」や、「接頭辞保存拡張技術」、「知識索引技術」等のビッグデータ処理の基盤技術・方式の研究開発を強力に進めてきた。
- ・特に、木構造や、グラフ、系列等の離散データに対する理論的性能保証付き極大パターン発見アルゴリズムについては、申請者らによる初めての定式化を行い、欧州の一級のデータマイニング研究グループ・研究者を中心に理論的な検証や発展が増加しつつある（2(2)項の引用等参照）。
- ・特別推進研究でのこれらの技術の啓蒙・発展の活動を通じて、これらのアルゴリズムの中核となるいくつかの基本アルゴリズムは、当該分野において欧米の研究者から良く引用される理論的な基盤技術として定着してきた。例として、申請者らによる木構造データの基本アルゴリズム「最右拡張法」の原著論文（SIAM DM 2002）は引用416件であり、高速アルゴリズム LCM は3つの関連論文で引用409件、半構造索引（LCP配列）の高速構築技術は引用293件に達する等（いずれも2013年5月 Google Scholar 調べ）、理論分野としては引用が多く、当該分野の論文で必ず引用される基本文献の一つとなってきた。これらは、同技術の当初の開発のみならず、特別推進研究によるその後の継続的な技術開発と発展に負うところが大きい。

#### ◎超高速極大集合発見アルゴリズム LCM の改良と発展：

- ・申請者らが開発し、本特別委推進研究で発展させた超高速極大集合発見アルゴリズム LCM（Linear-time Closed set Miner）は、列挙アルゴリズムの理論に基づく数理的アプローチにより、データマイニングの中心的課題の一つである頻出／極大パターンマイニングにおいて、最も高性能な方式の一つとして認識されている。
- ・同 LCM アルゴリズムは、第1版（2003年）および第2版（2004年）、特別推進研究による第3版（2005年）と改良され、各版の報告論文は、117件および、208件、84件の（排他的）引用を受け、合計409件の国際会議発表と学術論文から引用されている（2013年5月 Google Scholar 調べ）。また分担者の宇野による公開アルゴリズムもさまざまな応用に広く用いられている。
- ・同 LCM アルゴリズムは、2006年に IEEE データマイニング国際会議（ICDM2006）プログラム委員会が選定した「トップ10データマイニングアルゴリズム報告」（解説学術論文：Wu, Kumar, Quinlan 他, “Top 10 algorithms in data mining”, Knowledge and Information Systems, Vol.14）の中で、トップ10の一つ「頻出集合マイニングアルゴリズム」の項目の中で、特に「閉集合（極大）マイニング分野で最速のアルゴリズム」として言及され、「世界最速の閉集合マイニングアルゴリズム」として広く認知されるようになった。

#### ◎データマイニングと知識獲得研究分野の拡大のための活動：

- ・本特別推進研究では、知識基盤形成のための知識発見技術の目標として「高速かつ頑健な知識発見技術」を目標の一つに掲げて、研究を推進した。これに関して、半構造マイニングを含む大規模データからの知識獲得研究分野の拡大のためのさまざまな活動を行った。
- ・その一環として、本プロジェクトの主催で、データマイニングと数理統計分野の若手研究者を集めて2006年に「第1回データマイニングと統計数理国際ワークショップ」(The 1st International Workshop on Data-Mining and Statistical Science, DMSS2006, Sapporo, September 2006)を組織し、開催した。同国際ワークショップは、外部組織に移管しながら、第5回のDMSS2011まで開催されており（2012年移行はIBISワークショップの国際部門に統合）、ビッグデータ研究の隆盛以前の2000年代中期において、わが国におけるデータマイニングと知識獲得研究の拡大に微力ながら貢献した。

## 2. 特別推進研究の研究成果が他の研究者により活用された状況（続き）

(2) 論文引用状況（上位10報程度を記述してください。）

## 【研究期間中に発表した論文】

No	論文名	日本語による簡潔な内容紹介	引用数
1	Takeaki Uno, Masashi Kiyomi, Hiroki Arimura: LCM ver.3: Collaboration of Array, Bitmap and Prefix Tree for Frequent Itemset Mining, Proc. Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations (OSDM'05), ACM SIGKDD, 77-86, 2005.	申請者らが開発し、本特別委推進研究で発展させた超高速極大集合発見アルゴリズム LCM の第3版の実装報告。米国情報処理学会 ACM の知識発見国際会議 KDD2005 の併設ワークショップ OSDM2005 で発表された。同アルゴリズムは広く利用され、第1版の原著論文は現在 117 件、第2版論文は現在 208 件の学術論文・国際会議発表他から引用されている（2013年5月 Google Scholar 調べ）	84 件 （2013年5月 Google Scholar 調べ）
2	Hiroki Arimura, Takeaki Uno: An Output-Polynomial Time Algorithm for Mining Frequent Closed Attribute Trees, Lecture Notes in Computer Science 3625, Proc. 15th International Conference on Inductive Logic Programming (ILP'05), Springer, 1-19, 2005.	申請者らが開発したアルゴリズム FREQT を、本特別委推進研究において極大パターン発見技法を用いて高速化したアルゴリズム CLOATT を提案している。本特別推進研究の基盤技術の一つとして発展研究がおこなわれた FREQT アルゴリズム（原著論文 2002年）は日本発の世界初めての理論的性能保証付き木構造マイニングアルゴリズムとして、世界中で広く認知・利用され、第1版の原著論文は現在 416 件の学術論文・国際会議発表他から引用されている（2013年5月 Google Scholar 調べ）	30 件 （2013年5月 Google Scholar 調べ）
3	Hiroki Arimura, Takeaki Uno, An Efficient Polynomial Space and Polynomial Delay Algorithm for Enumeration of Maximal Motifs in a Sequence, Special issue on bioinformatics, Journal of Combinatorial Optimization, Vol.13, 243-262, 2006.	特別推進研究で開発した大規模系列データからの極大パターン発見アルゴリズム MAXMOTIF を提案している。本論文は、系列極大パターンの理論的性能保証付アルゴリズムに関する長年の未解決問題を肯定的に解決しており、2000年代末頃から欧州の研究者によって本論文が再度注目され、理論的研究が続いている(Esko Ukkonen, Theoretical Computer Science, Vol.410 (43), 2009 等)。	20 件 （2013年5月 Google Scholar 調べ）
4			
5			
6			
7			
8			
9			
10			

## 【研究期間終了後に発表した論文】

No	論文名	日本語による簡潔な内容紹介	引用数
1	H. Arimura and T. Uno: Polynomial-Delay and Polynomial-Space Algorithms for Mining Closed Sequences, Graphs, and Pictures in Accessible Set Systems, Proc. 2009 SIAM International Conference on Data Mining 2002 (SDM'09), SIAM, 1087-1098, 2009.	特別推進研究の研究成果をもとに構築された性能保証付極大半構造マイニングアルゴリズムの一般理論を、米国応用数学会 (SIAM) のデータマイニング国際会議で発表したもの。本論文は主に欧州のデータマイニング研究者に注目され (M. Boley, T. Horvath, A. Poigne, S. Wrobel, Theoretical computer science, Vol.411(3), 2010; S Nijssen, T Guns, PKDD2010 等)、また、大規模並列マイニングシステムの設計にも用いられている (Negrevergne, Termier 他, IEEE HPCS2010)。	19 件 (2013 年 5 月 Google Scholar 調べ)
2	Hiroki Arimura and Takeaki Uno: Mining Maximal Flexible Patterns in a Sequence, New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science 4914, Springer, 307-317, 2008.	特別推進研究で開発した大規模系列データからの伸び縮みを許した系列パターン発見アルゴリズム MAXFLEX を提案している。2000 年代末頃から欧州の情報科学と生物情報学の研究者らによって本論文が再度注目され、伸縮系列パターン発見についてピサ大の Robert Gross 教授、Nadia Pisanti らの研究グループでの理論的研究が続いている (Gross, Pisanti 他, Journal of Computational Biology, Vol.18(4), 2011 等)。	12 件 (2013 年 5 月 Google Scholar 調べ)
3			
4			
5			
6			
7			
8			
9			
10			



### 3. その他、効果・効用等の評価に関する情報

次の(1)、(2)の項目ごとに、該当する内容について具体的かつ明確に記述してください。

#### (1) 研究成果の社会への還元状況（社会への還元の程度、内容、実用化の有無は問いません。）

◎超高速極大集合発見アルゴリズム LCM（論文引用 2(2)[1]）。

申請者らが開発し、本特別委推進研究においてその改良版を研究開発した超高速極大集合発見アルゴリズム LCM のアルゴリズムの分担者の宇野による実装が公開され、世界中のデータマイニング研究者と実務家に広く用いられている。このアルゴリズムは、「世界最速の極大集合発見アルゴリズム」(“The world fastest closed pattern miner”) (IEEE ICDM2006, Program Committee) として広く知られている。申請者らが 2003 年に第 1 版を開発し、2004 年に第 2 版、2005 年に本特別推進研究において、第 3 版を研究開発したものであり、第 1 版の原著論文は現在 117 件、第 2 版論文は現在 208 件の学術論文・国際会議発表他から引用されており、現在も引用数が増加している。(2013 年 5 月 Google Scholar 調べ)

◎知識索引 ZDD と超高速極大集合発見手法 LCM の連携。

本特別推進研究では、超高速極大集合発見アルゴリズム LCM(ver3) および、それを分担者の湊が開発した大規模知識索引 ZDD 技術と結合した LCMoverZDD を研究開発した。この LCM および LCMoverZDD が、経産省「情報大航海プロジェクト」(2007-2009 年) のコラボレーションプラットフォームの「頻出パターンマイニング技術」として選定され、その実装が同コンソーシアムにおいて公開された。

URL: [http://www.meti.go.jp/policy/it\\_policy/daikoukai/igvp/cp\\_jp/](http://www.meti.go.jp/policy/it_policy/daikoukai/igvp/cp_jp/)

(同 HP から HOME > プロファイル情報解析層 > 頻出パターンマイニング技術)

**3. その他、効果・効用等の評価に関する情報（続き）****(2) 研究計画に関与した若手研究者の成長の状況（助教やポスドク等の研究終了後の動向を記述してください。）**

本研究メンバーには、助教とポスドクは含まれていない。本特別推進研究で立ち上げたテーマに関して、研究終了後の継続研究に参加した若手研究者（当時、博士後期課程の学生２名：2011年学位取得，2012年学位取得）が学位取得後に研究者として企業の研究職に就職した。

1名は、情報処理学会 平成22年度 IPSJ 論文船井若手奨励賞(2011年3月)を受賞した。